

The estimation of Wasserstein and Zolotarev distances to the class of exponential variables

Amparo Baíllo^a, Javier Cárcamo^a and Konstantin Getman^b

^a Departamento de Matemáticas, Universidad Autónoma de Madrid (Spain)

^b Department of Astronomy and Astrophysics, Pennsylvania State University (U.S.A.)

March 23, 2016

Abstract

Given a positive random variable X , we are interested in measuring how well the exponential distribution with the same mean approximates the probability distribution of X , based on the information provided by a sample from X . Specifically, we consider the problem of estimating the Zolotarev distance of order r , $r \in \mathbb{N}$, between X and the exponential distribution with mean $E(X)$. We give sharp results on the asymptotic distribution of a plug-in estimator of this metric and compare it with the finite-sample distribution via simulations. The practical use of the Zolotarev metrics is illustrated analysing a massive data set from X-ray astronomy.

Keywords: Zolotarev metric; plug-in estimator; asymptotic distribution; integrated Brownian bridge; integrated empirical process; probability metrics.

MSC: Primary 62E20; secondary 62G20.

1 Introduction

The exponential family is one of the most important and widely used parametric families of probability distributions in statistical applications. For instance, as exponential variables describe the lengths between consecutive occurrences of a homogeneous Poisson process, they are frequently used to model many physical phenomena. The mathematical simplicity of the exponential distribution usually allows to obtain explicit formulas and closed-form expressions (in terms of elementary functions) in those problems in which exponential variables are involved. Therefore, as it was already pointed out in [19], exponential models are frequently considered as a first-step approximation to other more complex models that could be more appropriate for a particular situation. For this reason, exponential distributions are often overused in applications. Statistical tools to evaluate

the error made in the approximation of a random variable by an exponential one are therefore worthwhile.

The exponential distribution also exhibits an uncountable number of significant mathematical properties (see the monographs by [5] and [1]) and many important characterizations (see, for instance, [3] and [13]). The class of exponential variables also plays a prominent role in reliability and life testing analysis (see [6], or, more recently, [21] and [24]). The exponential family is, undoubtedly, the most important one-parameter family of life distributions. Further, the most frequently used families of life distributions can be viewed as multi-parameter extensions of the exponential distribution.

Among the multiple characterizations of exponentiality, we highlight its lack of memory and constant hazard rate. Due to these two properties, the main classes of ageing distributions in reliability theory are usually defined by comparing some characteristic of a random variable with the corresponding one of an exponential variable of the same expectation. Other classes lifetime distributions are characterized by establishing an appropriate stochastic ordering relationship with respect to the exponential variable (of equal mean).

In this work, we follow a completely different approach. Instead of comparing a specific property of a random variable with the corresponding one of an exponential distribution, we aim to compute the distance of the random variable to the class of exponential ones. To be more precise, from now on we denote by Y_ν the random variable with exponential distribution with mean $\nu > 0$ and by $G_\nu(x) := 1 - \exp(-x/\nu)$, $x \geq 0$, its (cumulative) distribution function. Define \mathcal{E} as the *exponential class* given by

$$\mathcal{E} := \{Y_\nu : \nu \in (0, \infty)\}.$$

For a positive random variable X and an appropriate probability metric d , we are interested in estimating the distance

$$d(X, \mathcal{E}) := \inf\{d(X, Y_\nu) : \nu \in (0, \infty)\}. \quad (1)$$

We only consider *primary distances*, that is, metrics between probability measures. Hence, $d(X, Y)$ is completely determined by the pair of marginal distributions of X and Y . For notational convenience, if F and G are the distribution functions of X and Y , respectively, we indistinctly use $d(F, G)$ or $d(X, Y)$ and $d(F, \mathcal{E})$ or $d(X, \mathcal{E})$. We also restrict ourselves to *simple metrics*, i.e., metrics such that $d(F, G) = 0$ is equivalent to $F = G$.

For some of the metrics considered in this paper, it can be checked that $d(X, \mathcal{E}) = d(X, Y_\mu)$, where $\mu := EX$ is the expectation of X . However, in some other examples, it is not easy to determine the minimal element in \mathcal{E} for which the infimum in (1) is attained. In such cases, we simply propose to estimate $d(X, Y_\mu)$. In other words, although in general $d(X, \mathcal{E}) < d(X, Y_\mu)$, we consider Y_μ as the natural representant in \mathcal{E} to measure the distance of X to the exponential class.

Information on $d(X, Y_\mu)$ can be used to assert whether the exponential model provides a reasonable approximation of X . We also observe that, even within the same class of (ageing) distributions, two (lifetime) variables may have a very different behavior because they can be far away one from the other. Therefore, it may be interesting to quantify

the distance to the exponential class to have a better understanding of the underlying properties of the corresponding random variables.

The distance $d(X, Y_\mu)$, where X runs in a set of random variables \mathcal{X} , can also be used as a *clustering variable*, that is, a variable that detects groups or similitudes among the variables in \mathcal{X} . In this framework, we might further define collections of (integrable) distributions in terms of the proximity to the exponential class. For $\epsilon \geq 0$, we define the (ϵ, d) -neighbourhood of \mathcal{E} as

$$\mathcal{F}_d(\epsilon) := \{X : d(X, Y_\mu) \leq \epsilon\}. \quad (2)$$

We observe that $\mathcal{F}_d(0) = \mathcal{E}$ (the exponential class) and, for $\epsilon > 0$, $\mathcal{F}_d(\epsilon)$ can be viewed as the class of exponential random variables for which we allow a degree of contamination ϵ , measured by the distance d . In other words, $\mathcal{F}_d(\epsilon)$ is a full topological neighborhood of the exponential class. This definition follows the spirit of robustness by [18], where breakdown points are defined in terms of metrics, such as the Lévy, Kolmogorov and Prohorov metrics. Actually, in general $\mathcal{F}_d(\epsilon)$ contains neighborhoods of exponential variables constituted by epsilon-contaminated exponential mixtures (see Section 2.4).

In practice, the distribution of the random variable X is usually unknown. Therefore, the distance $d(X, Y_\mu)$ has to be estimated based a random sample X_1, \dots, X_n from X . We propose to use the plug-in estimator obtained by replacing the true distribution F of X by the empirical distribution F_n of the sample X_1, \dots, X_n , i.e.,

$$F_n(t) := \frac{1}{n} \sum_{i=1}^n I_{\{X_i \leq t\}}, \quad n \in \mathbb{N}, \quad t \geq 0,$$

where I_A stands for the indicator function of the set A . Thus, we consider $d(F_n, G_{\hat{\mu}})$ as the natural estimator of $d(F, G_\mu)$, where $\hat{\mu} = \frac{1}{n} \sum_{i=1}^n X_i$ is the sample mean.

In this work we are interested in the asymptotic behavior of $d(F_n, G_{\hat{\mu}})$. As it follows from general results in [27], under appropriate assumptions, it can be shown that $d(F_n, F) = O_P(1/\sqrt{n})$. Consequently, we want to determine the asymptotic distribution of

$$\delta_n(d, F) := \sqrt{n} (d(F_n, G_{\hat{\mu}}) - d(F, G_\mu)), \quad n \in \mathbb{N}. \quad (3)$$

A related problem was considered by [33], who obtained strong consistency and rates of convergence for a plug-in estimator of integral probability metrics between two probability measures.

This paper is structured as follows. In the next section, we present the Zolotarev ideal metrics, ζ_r ($r \in \mathbb{N}$), that we will discuss throughout this work. These metrics are suitable for the problem under consideration as $\zeta_r(X, Y_\mu)$ measures the maximum error made in the approximation of X by Y_μ in expected value within a certain class of smooth functions. Moreover, the metrics ζ_r are more sensitive to extreme values than the usual Kolmogorov metric. Among Zolotarev metrics, we check that the only possible choices are the so-called Wasserstein distance and the Zolotarev ζ_2 -metric. The Wasserstein metric has found applications in many areas (see Section 2.1), whereas, as mentioned in [28, Section 15], the Zolotarev ζ_2 -metric is appropriate for investigating some ageing properties of lifetime distributions. We also enumerate some properties of these two distances, recall

their dual representations, and provide explicit expressions for $\delta_n(d, F)$. As these two metrics are not scale invariant, we propose to use normalized versions of the distances to obtain dimensionless quantities. For these distances, the relationship between $\mathcal{F}_d(\epsilon)$ and the usual ϵ -contamination (by mixtures) neighborhoods of exponential variables is also briefly discussed. In Section 3, we determine the asymptotic distribution of $\delta_n(d, F)$, for the Wasserstein distance, the Zolotarev ζ_2 -metric, and their normalized counterparts. We also obtain sharp asymptotic results on the convergence of the underlying stochastic process in the space $L^1 \equiv L^1(0, \infty)$ (endowed with the usual norm $\|\cdot\|_1$), which can be of independent theoretical interest in other problems related to the exponential distribution. In Section 4, a simulation study is carried out to assess the practical performance of our proposal with finite samples. A real data set is analyzed in Section 5. Finally, detailed proofs of the main results are included in the technical appendix.

2 Zolotarev ideal probability metrics

As stated in Section 1 we want to measure how far the probability distribution of a positive random variable X of interest is from the exponential class \mathcal{E} . A central issue is, thus, to choose a suitable metric quantifying such a distance. Throughout this section, X and Y denote two general positive random variables with cumulative distribution functions F and G , respectively. When we assume that Y follows an exponential distribution with mean μ , then we specifically denote it by Y_μ and its distribution function by G_μ .

In this work, we consider the *Zolotarev metric* of order r , $r \in \mathbb{N}$, (see [35]), defined by

$$\zeta_r(X, Y) := \sup_{f \in \mathcal{F}_r} |Ef(X) - Ef(Y)|, \quad (4)$$

where \mathcal{F}_r is the class of $(r - 1)$ -times continuously differentiable functions $f : \mathbb{R} \rightarrow \mathbb{R}$ satisfying the Lipschitz condition $|f^{(r-1)}(x) - f^{(r-1)}(y)| \leq |x - y|$, for all $(x, y) \in \mathbb{R}^2$. We use the notation $f^{(0)} \equiv f$. The class \mathcal{F}_r can also be substituted by the set of functions f having r -th derivative $f^{(r)}$ a.e. and such that $|f^{(r)}| \leq 1$ a.e. For a general reference and properties of the distance ζ_r we refer to [28].

The metric ζ_r is *ideal of order r* , that is, it satisfies the following two properties:

- (a) *Regularity*: $\zeta_r(X + Z, Y + Z) \leq \zeta_r(X, Y)$, for any random variable Z , independent of the vector (X, Y) .
- (b) *Homogeneity of order r* : $\zeta_r(cX, cY) = |c|^r \zeta_r(X, Y)$, for all $c \in \mathbb{R}$.

The metric ζ_r is important because it metrizes weak convergence plus convergence of r -th absolute moments. For this reason, and because ζ_r is ideal, these metrics appear naturally when studying problems related to limit theorems in probability theory (see [29]). The case $r = 1$ has special interest, as ζ_1 coincides with the so-called Wasserstein distance. The metric ζ_2 has been considered in [28, Section 15] to check the robustness of a χ^2 -test of exponentiality. The metric ζ_3 has been applied in the context of distributional recurrences in [25] and [26]. Further, several bounds involving Zolotarev metrics can be given for other probability metrics such as Kolmogorov and Lévy metrics (see [28]).

Definition (4) shows that $\zeta_r(X, Y_\mu)$ represents the maximum error in the expected value within the class of functions \mathcal{F}_r due to the approximation of X by Y_μ . In particular, for every function f with r -th derivative, we have that

$$|Ef(X) - Ef(Y_\mu)| \leq \zeta_r(X, Y_\mu) \|f^{(r)}\|_\infty,$$

where $\|\cdot\|_\infty$ is the sup-norm.

The finiteness of $\zeta_r(X, Y)$ guarantees the equality of all moments of X and Y up to order r (i.e., $EX^k = EY^k$, for $k \in \{1, \dots, r-1\}$). Conversely, if $EX^k = EY^k$ for $k \in \{1, \dots, r-1\}$ and $E|X|^r, E|Y|^r < \infty$, then $\zeta_r(X, Y)$ is also finite. In particular, we see that if X is a positive random variable with mean $\mu > 0$, then

$$\zeta_r(X, \mathcal{E}) = \zeta_r(X, Y_\mu), \quad \text{for } r \geq 2. \quad (5)$$

However, in general, when $r \geq 3$ and X is *not* exponential, we have that $\zeta_r(X, Y_\mu) = \infty$. This follows from the fact that, for many distributions, equalities $EX = \mu$ and $EX^2 = 2\mu^2$ are too restrictive and they actually imply that X is exponential. For instance, this happens for the classes HNBUE or HNWUE variables (see Section 2.2 for more details on these classes of ageing distributions), as it can be checked by known results on stochastic equality under convex domination (see for instance [32, Theorem 3.A.42, p. 133]). Therefore, only the cases $r = 1$ and $r = 2$ make sense for the discussed problem.

2.1 The Wasserstein distance

The case $r = 1$ in (4) has special relevance. By the Kantorovich–Rubinstein theorem, we see that $\zeta_1 \equiv \omega$, where ω is the famous *L^1 -Wasserstein metric*, also known as the *Kantorovich–Rubinstein distance*. In the context of image processing, Wasserstein distance is known as the *earth mover's distance* (see [30]). [17] and [28] collect some relationships and bounds between the Wasserstein metric and other important probability metrics, such as the Prokhorov, the discrepancy, and the total variation metrics. For this distance, in general $\omega(F, \mathcal{E}) < \omega(F, G_\mu)$.

It is well-known that $\omega(F, G)$ can be also expressed as

$$\omega(F, G) = \int_0^\infty |F(t) - G(t)| dt. \quad (6)$$

Using (6), we obtain that

$$\delta_n(\omega, F) = \sqrt{n} \left(\int_0^\infty |F_n(t) - G_{\hat{\mu}}(t)| dt - \int_0^\infty |F(t) - G_\mu(t)| dt \right).$$

2.2 The Zolotarev ζ_2 -metric

As argued in [28, p. 340], the Zolotarev ζ_2 -metric is the “natural metric” when dealing with the exponential class.

If $\zeta_2(X, Y)$ is finite, then $EX = EY$. In that case, ζ_2 admits the dual representation

$$\zeta_2(X, Y) = \int_0^\infty \left| \int_t^\infty (F(x) - G(x)) dx \right| dt. \quad (7)$$

From (7), we have the bounds

$$|EX^2 - EY^2| \leq 2\zeta_2(X, Y) \leq EX^2 + EY^2. \quad (8)$$

In risk analysis, $\zeta_2(X, Y)$ is called the *integrated stop-loss distance* (see [10]) because, by Fubini's theorem, it can be readily checked that

$$\zeta_2(X, Y) = \int_0^\infty |E(X - t)_+ - E(Y - t)_+| dt,$$

where $(a)_+ := \max\{a, 0\}$ is the positive part of the real number a and $E(X - t)_+$ ($t \geq 0$) is the so-called *stop-loss function* of X .

Remark 1. The metrics ω and ζ_2 are more sensitive to the differences in the probabilities corresponding to extreme values than other usual probability metrics such as the *Kolmogorov distance* $\kappa(X, Y) := \sup_x |F(x) - G(x)|$. As the difference $|F(x) - G(x)|$ converges to zero as x increases or decreases, the contribution of the terms corresponding to extreme events is usually small. As a consequence, the differences in the tail behavior of X and Y will only be reflected in $\kappa(X, Y)$ to a relatively small extent. However, the representations (6) and (7) show that extreme values have more weight in ω and ζ_2 as integrals of tail probabilities appear in these distances.

If X has finite expectation $\mu > 0$, by (5) and (7) we have that

$$\zeta_2(F, \mathcal{E}) = \zeta_2(X, Y_\mu) = \int_0^\infty \left| \int_t^\infty (F(x) - G_\mu(x)) dx \right| dt \quad (9)$$

and

$$\zeta_2(F_n, \mathcal{E}) = \zeta_2(F_n, G_{\hat{\mu}}) = \int_0^\infty \left| \int_t^\infty (F_n(x) - G_{\hat{\mu}}(x)) dx \right| dt.$$

In the context of life distributions (see [24]), a (nonnegative) random variable X with distribution function F is said to be *harmonic new better than used in expectation* (HNBUE) if

$$\int_t^\infty (1 - F(x)) dx \leq \mu \exp(-t/\mu), \quad \text{for all } t \geq 0. \quad (10)$$

Analogously, X is *harmonic new worse than used in expectation* (HNWUE) if the reverse inequality in (10) holds. The HNBUE (respectively, HNWUE) class is fairly large and includes all the usual ageing (respectively, anti-ageing) classes such as IFR, IFRA, NBU, and NBUE (respectively, DFR, DFRA, NWU, and NWUE) (see [6] for the precise definitions of these classes). From (9) and (10), whenever X belongs to the class HNBUE (or HNWUE with finite second moment), we actually have that

$$\zeta_2(X, \mathcal{E}) = \frac{1}{2} |\mu^2 - \sigma^2|, \quad (11)$$

where σ^2 is the variance of X (see [8],[28, Section 15]).

2.3 Normalized Wasserstein and Zolotarev metrics

The homogeneity property of ζ_r , $r = 1, 2$, implies that, if X is measured in certain units, then $\zeta_r(X, Y_\mu)$ is expressed in units ^{r} . To obtain adimensional distances, we consider the following modification of ω and ζ_2 for random variables with finite and positive expectation:

$$\bar{\omega}(X, Y) := \omega(X/\mathbb{E}X, Y/\mathbb{E}Y) \quad \text{and} \quad \bar{\zeta}_2(X, Y) := \zeta_2(X/\mathbb{E}X, Y/\mathbb{E}Y).$$

We call $\bar{\omega}$ and $\bar{\zeta}_2$ the *normalized Wasserstein* and *normalized Zolotarev ζ_2 -metrics*, respectively. Obviously, these normalized metrics are homogeneous of degree 0 and scale independent (they have no units). It is also straightforward to check that, if $\mathbb{E}X = \mathbb{E}Y = \mu$, then

$$\bar{\omega}(X, Y_\mu) = \frac{1}{\mu} \omega(X, Y_\mu) \quad \text{and} \quad \bar{\zeta}_2(X, \mathcal{E}) = \frac{1}{\mu^2} \zeta_2(X, Y_\mu).$$

From (8), for a variable X with finite second moment, we obtain the bounds

$$|\text{CV}(X)^2 - 1| \leq 2 \bar{\zeta}_2(X, \mathcal{E}) \leq 3 + \text{CV}(X)^2, \quad (12)$$

where $\text{CV}(X) := \sigma/\mu$ is the coefficient of variation of X . Observe that the first inequality in (12) is actually an equality if X belongs to the class HNBUE or HNWUE.

2.4 Contaminated exponential variables

As mentioned in the introduction, $\mathcal{F}_d(\epsilon)$, the (ϵ, d) -neighbourhood of \mathcal{E} defined in (2), can be considered as the class of “almost exponential” variables for which we allow a degree ϵ of contamination measured by d . Let us see that, for the considered distances, $\mathcal{F}_d(\epsilon)$ is much larger than the usual ϵ -contamination neighborhoods of exponential variables obtained by mixtures.

Consider an exponential variable Y_λ contaminated by mixing it with a positive random variable Z (with distribution function H and $\mathbb{E}Z = \nu > 0$). In other words, for $\epsilon \in [0, 1]$, define the random variable X_ϵ with distribution function $(1 - \epsilon)G_\lambda + \epsilon H$. The quantity ϵ , usually a small number below 0.5, is the *fraction of contamination* added to Y_λ .

For the metric ω , setting $\mu := \epsilon \nu + (1 - \epsilon)\lambda$ (the expectation of X_ϵ), we have that

$$\begin{aligned} \omega(X_\epsilon, Y_\mu) &= \|(1 - \epsilon)G_\lambda + \epsilon H - G_\mu\|_1 \\ &\leq \|G_\lambda - G_\mu\|_1 + \epsilon(\|G_\nu - G_\lambda\|_1 + \|H - G_\nu\|_1) \\ &= \epsilon(2|\lambda - \nu| + \omega(H, G_\nu)), \end{aligned}$$

In particular, we obtain that $\omega(X_\epsilon, Y_\mu) \leq 2(\nu + |\lambda - \nu|)\epsilon$, and we see that $X_\epsilon \in \mathcal{F}_\omega(M\epsilon)$, with $M := 2(\nu + |\lambda - \nu|)$.

For ζ_2 , similar computations show that

$$\zeta_2(X_\epsilon, \mathcal{E}) \leq \epsilon(|\lambda - \nu|(3\lambda + \nu + \epsilon(\nu - \lambda)) + \zeta_2(Z, \mathcal{E})).$$

Therefore, we always have that $X_\epsilon \in \mathcal{F}_{\zeta_2}(N\epsilon)$, with $N := |\lambda - \nu|(3\lambda + \nu + \epsilon(\nu - \lambda)) + \mathbb{E}Z^2/2 + \nu^2$.

3 Asymptotic behavior of the estimators of the distances

In this section, we determine the asymptotic distribution of $\delta_n(d, F)$ in (3), for $d = \omega, \bar{\omega}, \zeta_2$ and $\bar{\zeta}_2$. Though the detailed proofs are collected in the Appendix, we describe here in broad strokes the main ideas behind them. First, we note that

$$\delta_n(d, F) = \rho_n(\mathbb{X}_n(d), g_d), \quad (13)$$

where $\rho_n : L^1 \times L^1 \rightarrow \mathbb{R}$ is the linear functional defined by

$$\rho_n(f, g) := \|f + \sqrt{n}g\|_1 - \sqrt{n}\|g\|_1, \quad \text{for } f, g \in L^1, \quad (14)$$

$\mathbb{X}_n(d)$ are the stochastic processes given (for $t \geq 0$) by

$$\begin{aligned} \mathbb{X}_n(\omega, t) &:= \sqrt{n} [(F_n(t) - G_{\hat{\mu}}(t)) - (F(t) - G_{\mu}(t))], \\ \mathbb{X}_n(\bar{\omega}, t) &:= \sqrt{n} \left[\frac{1}{\hat{\mu}} (F_n(t) - G_{\hat{\mu}}(t)) - \frac{1}{\mu} (F(t) - G_{\mu}(t)) \right], \\ \mathbb{X}_n(\zeta_2, t) &:= \sqrt{n} \left[\int_t^\infty (F_n(x) - G_{\hat{\mu}}(x)) dx - \int_t^\infty (F(x) - G_{\mu}(x)) dx \right], \\ \mathbb{X}_n(\bar{\zeta}_2, t) &:= \sqrt{n} \left[\frac{1}{\hat{\mu}^2} \int_t^\infty (F_n(x) - G_{\hat{\mu}}(x)) dx - \frac{1}{\mu^2} \int_t^\infty (F(x) - G_{\mu}(x)) dx \right], \end{aligned} \quad (15)$$

and g_d are the (deterministic) functions defined by

$$g_\omega(t) := F(t) - G_{\mu}(t), \quad g_{\bar{\omega}}(t) := \frac{1}{\mu} (F(t) - G_{\mu}(t)), \quad (16)$$

$$g_{\zeta_2}(t) := \int_t^\infty (F(x) - G_{\mu}(x)) dx, \quad g_{\bar{\zeta}_2}(t) := \frac{1}{\mu^2} \int_t^\infty (F(x) - G_{\mu}(x)) dx. \quad (17)$$

From (13), we see that establishing the (weak) convergence in L^1 of the processes $\mathbb{X}_n(d)$ in (15), joint with the continuity of the linking functional in (14), immediately translates into the convergence in distribution of $\delta_n(d, F)$.

Before stating the main results, we need to introduce some definitions and notation. In the sequel, $\mathbb{B}_F := \mathbb{B} \circ F$ is the *F-Brownian bridge*, where \mathbb{B} is a standard Brownian bridge on $[0, 1]$, that is, \mathbb{B} is a centered Gaussian process with covariance function $\gamma(s, t) = s \wedge t - st$ and continuous paths, with probability 1.

We consider the *Lorentz spaces* of positive random variables defined by

$$\mathcal{L}^{2,1} := \{X : \Lambda_{2,1}(X) < \infty\} \quad \text{and} \quad \mathcal{L}^{4,2} := \{X : \Lambda_{4,2}(X) < \infty\},$$

where

$$\Lambda_{2,1}(X) := \int_0^\infty \sqrt{\mathbb{P}(X > t)} dt \quad \text{and} \quad \Lambda_{4,2}(X) := \int_0^\infty t \sqrt{\mathbb{P}(X > t)} dt$$

(see [22, p. 279]). Conditions $\Lambda_{2,1}(X) < \infty$ and $\Lambda_{4,2}(X) < \infty$ are slightly stronger than $\mathbb{E}X^2 < \infty$ and $\mathbb{E}X^4 < \infty$, respectively (see [12]). It is known (see [7]) that $X \in \mathcal{L}^{2,1}$ is

equivalent to the convergence of the empirical process to \mathbb{B}_F in L^1 . Finally, $\mathcal{L}^p := \{X : EX^p < \infty\}$ ($p > 0$) is the usual \mathcal{L}^p space of random variables with finite p -th moment.

The following two theorems, which are of independent theoretical interest by themselves, characterize the asymptotic behavior of $\mathbb{X}_n(d)$ in L^1 . The results are sharp in the sense that we obtain the exact integrability condition on X so that the processes converge in distribution in L^1 . The symbol " $\xrightarrow{L^1}_w$ " stands for the weak convergence of a sequence of random processes in the space L^1 (see the Appendix for the precise definition).

Theorem 1. *Let X be a positive random variable with expectation $\mu > 0$. If $X \in \mathcal{L}^{4/3}$, the following assertions are equivalent:*

(a) $X \in \mathcal{L}^{2,1}$.

(b) $\mathbb{X}_n(\omega) \xrightarrow{L^1}_w \mathbb{X}_F(\omega)$, where $\mathbb{X}_F(\omega)$ is a centered Gaussian process given by

$$\mathbb{X}_F(\omega, t) := \mathbb{B}_F(t) - \frac{t}{\mu^2} e^{-t/\mu} \int_0^\infty \mathbb{B}_F, \quad t \geq 0. \quad (18)$$

(c) $\mathbb{X}_n(\bar{\omega}) \xrightarrow{L^1}_w \mathbb{X}_F(\bar{\omega})$, where $\mathbb{X}_F(\bar{\omega})$ is a centered Gaussian process given by

$$\mathbb{X}_F(\bar{\omega}, t) := \frac{1}{\mu} \left[\mathbb{B}_F(t) + \left(g_{\bar{\omega}}(t) - \frac{t}{\mu^2} e^{-t/\mu} \right) \int_0^\infty \mathbb{B}_F \right], \quad t \geq 0, \quad (19)$$

and the function $g_{\bar{\omega}}$ is defined in (16).

Theorem 2. *Let X be a positive random variable with expectation $\mu > 0$. The following assertions are equivalent:*

(a) $X \in \mathcal{L}^{4,2}$.

(b) $\mathbb{X}_n(\zeta_2) \xrightarrow{L^1}_w \mathbb{X}_F(\zeta_2)$, where $\mathbb{X}_F(\zeta_2)$ is a centered Gaussian process given by

$$\mathbb{X}_F(\zeta_2, t) := \int_t^\infty \mathbb{B}_F - \left(1 + \frac{t}{\mu} \right) e^{-t/\mu} \int_0^\infty \mathbb{B}_F, \quad t \geq 0. \quad (20)$$

(c) $\mathbb{X}_n(\bar{\zeta}_2) \xrightarrow{L^1}_w \mathbb{X}_F(\bar{\zeta}_2)$, where $\mathbb{X}_F(\bar{\zeta}_2)$ is a centered Gaussian process given by

$$\mathbb{X}_F(\bar{\zeta}_2, t) := \frac{1}{\mu^2} \left[\int_t^\infty \mathbb{B}_F + \left(2\mu g_{\bar{\zeta}_2}(t) - \left(1 + \frac{t}{\mu} \right) e^{-t/\mu} \right) \int_0^\infty \mathbb{B}_F \right], \quad t \geq 0, \quad (21)$$

and the function $g_{\bar{\zeta}_2}$ is defined in (17).

Taking into account Theorems 1 and 2, to derive the asymptotic distribution of $\delta_n(d, F)$ it will suffice to analyze the continuity of ρ_n in (14). This is carried out in the next key lemma in which $\text{sgn}(\cdot)$ denotes the sign function and A^c is the complementary of the set A .

Lemma 1. For $f, g \in L^1$, the functional ρ_n defined in (14) satisfies that

$$\lim_{n \rightarrow \infty} \rho_n(f, g) = \rho(f, g) := \int_{I_g} |f| + \int_{I_g^c} f \operatorname{sgn}(g), \quad (22)$$

where $I_g := \{t : g(t) = 0\}$. Moreover, if $f_n \rightarrow f$ in L^1 , then $\rho_n(f_n, g) \rightarrow \rho(f, g)$.

Using (13), Theorems 1 and 2 and Lemma 1, we obtain the asymptotic distribution of $\delta_n(d, F)$. In the following theorem, we use the notation “ \rightarrow_d ” for the usual convergence in distribution of random variables.

Theorem 3. Let X be a positive random variable with expectation $\mu > 0$. For $d = \omega$ or $d = \bar{\omega}$ (respectively, for $d = \zeta_2$ or $d = \bar{\zeta}_2$), let us assume that $X \in \mathcal{L}^{2,1}$ (respectively, $X \in \mathcal{L}^{4,2}$). Then $\delta_n(d, F) \rightarrow_d \rho(\mathbb{X}_F(d), g_d)$, where ρ is defined in (22), the processes $\mathbb{X}_F(d)$ are defined in (18)-(21), and the functions g_d are given in (16)-(17).

The next corollary, a direct consequence of Theorem 3, gives the asymptotic distribution of $\delta_n(d, F)$, when F is an exponential distribution function.

Corollary 1. For the processes $\mathbb{X}_F(d)$ defined in (18)-(21), if X follows an exponential distribution with mean μ , then

- (a) $\sqrt{n} \omega(F_n, G_{\hat{\mu}}) \rightarrow_d \|\mathbb{X}_{G_{\mu}}(\omega)\|_1 = \mu \|\mathbb{X}_{G_1}(\omega)\|_1;$
- (b) $\sqrt{n} \bar{\omega}(F_n, G_{\hat{\mu}}) \rightarrow_d \|\mathbb{X}_{G_1}(\bar{\omega})\|_1 = \|\mathbb{X}_{G_1}(\omega)\|_1;$
- (c) $\sqrt{n} \zeta_2(F_n, G_{\hat{\mu}}) \rightarrow_d \|\mathbb{X}_{G_{\mu}}(\zeta_2)\|_1 = \mu^2 \|\mathbb{X}_{G_1}(\zeta_2)\|_1;$
- (d) $\sqrt{n} \bar{\zeta}_2(F_n, G_{\hat{\mu}}) \rightarrow_d \|\mathbb{X}_{G_1}(\bar{\zeta}_2)\|_1 = \|\mathbb{X}_{G_1}(\zeta_2)\|_1.$

In the following corollary, we show that when X does not share any part of its distribution function with the exponential one, $\delta_n(d, F)$ is actually asymptotically normal, for all the considered distances.

Corollary 2. Let us assume that the conditions of Theorem 3 hold and let us further assume that the set I_{g_d} (defined in Lemma 1) has zero Lebesgue measure in \mathbb{R} . Then the statistics $\delta_n(d, F)$ converge in distribution to a zero mean normal random variable.

4 A simulation study

The aim of this section is to compare the finite-sample distribution of the statistic $\delta_n(d, F)$ given in (3), for $d = \bar{\omega}$ and $d = \bar{\zeta}_2$, with its asymptotic distribution obtained in Theorem 3.

Computing the normalized version, $d = \bar{\omega}$ or $d = \bar{\zeta}_2$, of the distances $d(F, G_{\mu})$ is equivalent to computing the distance of the re-scaled variable X/μ to the exponential distribution with mean $\mu = 1$. As a consequence, in this Monte Carlo study the data-generating distributions were all taken with expectation $\mu = 1$. Specifically, we have considered

- the exponential distribution with mean $\mu = 1$;

- the Weibull distribution with shape parameter $a > 0$ and scale parameter $\lambda = 1/\Gamma(1 + 1/a)$, with probability density

$$f(x) = \frac{a}{\lambda} \left(\frac{x}{\lambda}\right)^{a-1} e^{-(x/\lambda)^a}, \quad \text{for } x > 0;$$

- the gamma distribution with shape parameter $a > 0$ and scale parameter $\lambda = 1/a$, with density

$$f(x) = \frac{a^a}{\Gamma(a)} x^{a-1} e^{-ax}, \quad \text{for } x > 0;$$

For the Weibull distribution with mean 1 we have chosen the values $a = 0.9$ and $a = 1.1$ for the shape parameter. For the gamma distribution with mean 1, we have used shape parameters $a = 0.9$, $a = 1.1$ and $a = 1.2$ (see Figure 1).

In the simulations we have taken 10000 samples of size $n = 100, 500, 1000$ and 5000 from each of these distributions. For each Monte Carlo sample we have computed the statistic $\delta_n(d, F)$ for $d = \bar{\omega}$ and $d = \bar{\zeta}_2$. The results of the simulations are summarized in 2, 5, 6, 7, 4 and 3. Each figure shows the evolution, as n increases, of the boxplots of these statistics $\delta_n(d, F)$ towards the boxplot of its asymptotic distribution. This latter boxplot is also based on 10000 samples from the corresponding limit distribution.

On the one hand, we observe that, the finite-sample behaviour of $\delta_n(d, F)$, both for $d = \bar{\omega}$ and for $d = \bar{\zeta}$, is quite stable for the exponential distribution. In particular, the quartiles and median of $\delta_n(d, F)$ are very similar for any sample size. For distributions F other than the exponential, the boxplot of $\delta_n(d, F)$ resembles that of the limit distribution for large sample sizes ($n \geq 1000$). This is a typical behaviour of nonparametric function estimators such as the empirical distribution function, F_n . In a sense, it is the price to pay for estimating the distance between probability distributions (and not just summaries of these) and using an estimator of F which does not impose any parametric restrictions. On the other hand, we also observe that, the closer is F to the exponential distribution, the bigger n has to be for the distribution of $\delta_n(d, F)$ to approach its limit. For the examples considered, this fact is especially apparent if we compare the two gamma cases with shape parameters $a = 1.1$ and $a = 1.2$. This is a reasonable behaviour: when F is not exponential, but close to it and the set I_{g_d} has zero Lebesgue measure, the finite sample behaviour of $\delta_n(d, F)$ is almost as if F were exponential, but the limit actually has to be Gaussian (see Corollary 2).

5 Analysis of a real dataset: Chandra Orion Ultradeep Project

The aim of this section is to illustrate the use of the distance estimator $d(F_n, G_{\hat{\mu}})$ in practice, with a real dataset obtained as a result of Chandra Orion Ultradeep Project (COUP), a ground-breaking mission in X-ray astronomy. Among other things, the COUP studies X-flaring in pre-main-sequence (PMS) stars, members of the Orion Nebula region that is composed of the rich revealed Orion Nebula Cluster (ONC) and the filamentary molecular cloud called Orion Molecular Cloud 1 (OMC-1). With respect to the observer, the ONC cluster lies right in front of the OMC-1 cloud.

A PMS star is a premature star that has acquired all of its mass from its natal envelope of interstellar dust and gas and contracts until it starts hydrogen burning (see, e.g., [31]).

These young stars have intense magnetic fields, detected through their X-ray emissions, where plasma, confined in magnetic loops, is heated to X-ray emitting temperatures. Detection of these high-energy emissions is only possible by space observatories, such as the Chandra X-ray Observatory, an Earth satellite in a 64-hour orbit (see <http://chandra.si.edu/>). It is better to study the magnetic activity of young stars in large samples of these, due to the large variability of their X-ray emission levels. The nearest rich and concentrated collection of PMS stars is in the ONC/OMC-1 star forming region.

In January 2003 the Chandra X-ray Observatory focussed its Advanced CCD Imaging Spectrometer on the ONC for a period of 13.2 days (see [16]). The results were an almost continuous observation of the photon arrival times and associated energies for 1616 X-ray sources. COUP sources were compared with source positions from previously existing catalogs, with the aim of physically associating the COUP sources with already identified stars (whenever this was possible). The majority of these COUP sources has been classified into one of three groups (see [11]):

- Lightly-obscured PMS sources: This class is constituted by 835 cool low-mass PMS stars that are likely located in the ONC cluster. The term “lightly obscured” means that the star X-ray emission is less absorbed by the material in the interstellar medium.
- Heavily-obscured PMS sources: This class corresponds to 559 low-mass PMS stellar objects that are likely still embedded in the nascent OMC-1 cloud.
- Nonmembers: This group contains over 200 probable nonmembers of the Orion Nebula star forming region. A large part are extragalactic sources and a few are foreground stars or very faint sources without counterparts. For our analysis in this group we only consider the extragalactic X-ray emissions.

For each of the sources in these three groups we have computed the series of times between consecutive photon detections. Since for each COUP source we have a sample of the random variable of interest, namely, photon interarrival time (PIT), it is natural to compare different sources by computing the distances between their corresponding PITs and an exponential variable with the same mean. Extragalactic radiation does not exhibit flares and has usually a constant photon emission rate. Photon arrival times in this type of sources are well-modeled by a homogeneous Poisson process and, thus, their PITs follow an exponential distribution.

In order to get reasonable estimates of the distances to the exponential distribution, we kept only the COUP series with at least 100 photon inter-arrival times. We finally analyzed 1090 samples of PITs, of which 73, 644 and 373 were extragalactic sources, lightly-obscured and heavily-obscured PMS stars respectively. For each of these COUP sources, we compute the distance, $d(F_n, G_{\hat{\mu}})$, of the empirical distribution of PIT to the exponential distribution with the same sample mean. Here we have only used the scale invariant distances: the normalized Wasserstein metric $d = \bar{\omega}$ and the normalized Zolotarev metric $d = \bar{\zeta}_2$.

The results are summarized in Figures 8 and 9. An interesting issue is whether the distance of PIT to the exponential distribution depends on the source class. In Figure 8 we have

displayed the boxplots of $\log(d(F_n, G_{\hat{\mu}}))$, for $d = \bar{\omega}$ and $d = \bar{\zeta}_2$, separated according to the three types of COUP sources. In Figure 9 we plot the empirical distribution functions of $\sqrt{n}d(F_n, G_{\hat{\mu}})$ for the different COUP groups and the asymptotic distribution function of $\delta_n(d, G_1)$ as given by Corollary 1. As expected, we can see that the distance of interarrival photon times due to extragalactic radiation is the nearest to the exponential distribution. We also note that the distribution of interarrival times corresponding to lightly-obscured PMS stars is nearer to the exponential than that of the heavily-obscured group. This is reasonable, since X-ray emission is more affected by the intervening interstellar absorption, mainly the absorption from the gas in the molecular cloud and/or local absorption in an envelope or a disk around a star. Other factors might be also considered: since the heavily absorbed sample is generally younger and more “diskier” (the heavily absorbed sample has a higher fraction of stars that are still surrounded by their circumstellar disks; some of them highly accreting). Thus, the age and the presence/absence of disks could play an additional role, by affecting the X-ray production mechanisms of the strong X-ray flares in PMS stars and in turn affecting their X-ray photon arrival patterns (see [14] and [15] for differences in the strong X-ray flares between the diskless and disky-accreting stellar populations).

Appendix

In this technical appendix, we collect the proofs of the results stated in Section 3. The main ingredients are the following: first, we show that the considered sequences of stochastic processes are equivalent in L^1 to continuous functionals of the empirical process; then we use the central limit theorem (CLT) in suitable Banach spaces to find their weak limits; finally, we show the continuity of the mentioned functionals to derive the asymptotic distribution of $\delta_n(d, F)$.

First, we recall that if the stochastic processes \mathbb{P}_n and \mathbb{P} take values in L^1 a.s., it is said that \mathbb{P}_n converges in distribution to \mathbb{P} in L^1 if $\lim_{n \rightarrow \infty} \mathbb{E}f(\mathbb{P}_n) = \mathbb{E}f(\mathbb{P})$, for all continuous and bounded functions $f : L^1 \rightarrow \mathbb{R}$. Note that if \mathbb{P} and \mathbb{P}_n are jointly measurable and have almost all their trajectories in L^1 , they can be identified with Borel-measurable random elements in L^1 (see [9]). Therefore, the previous expectations are well-defined. In the following, we denote this weak convergence of probability measures in L^1 by $\mathbb{P}_n \xrightarrow{L^1}_w \mathbb{P}$. An analogous definition can be given for the weak convergence on other Banach spaces. Hence, from now on “ \xrightarrow{B}_w ” stands for the weak convergence of a sequence of random processes in the space B .

Let \mathbb{P}_n and $\tilde{\mathbb{P}}_n$ be two stochastic processes with trajectories in L^1 a.s. We say that \mathbb{P}_n and $\tilde{\mathbb{P}}_n$ are equivalent in L^1 , denoted $\mathbb{P}_n \stackrel{L^1}{\sim} \tilde{\mathbb{P}}_n$, if $\|\mathbb{P}_n - \tilde{\mathbb{P}}_n\|_1 \xrightarrow{\mathbb{P}} 0$, where “ $\xrightarrow{\mathbb{P}}$ ” stands for convergence in probability. Roughly speaking, if $\mathbb{P}_n \stackrel{L^1}{\sim} \tilde{\mathbb{P}}_n$, the two processes have the same asymptotic behavior in L^1 because if $\mathbb{P}_n \xrightarrow{L^1}_w \mathbb{P}$ and $\mathbb{P}_n \stackrel{L^1}{\sim} \tilde{\mathbb{P}}_n$, then $\tilde{\mathbb{P}}_n \xrightarrow{L^1}_w \mathbb{P}$ (see for instance [34, Theorem 18.10]).

In the sequel, \mathbb{E}_n stands for the *empirical process* associated to X , that is

$$\mathbb{E}_n(t) := \sqrt{n}(F_n(t) - F(t)), \quad t \geq 0, \quad n \geq 1.$$

The asymptotic behavior of \mathbb{E}_n in L^1 and in a weighted L^1 space are collected in the following lemma. Part (a) is a known result (see [7, Theorem 2.1]), while part (b) can be found in [4, Lemma 2].

Lemma 2. *We have that*

(a) $\mathbb{E}_n \xrightarrow{L^1}_w \mathbb{B}_F$ if and only if $X \in \mathcal{L}^{2,1}$.

(b) $\mathbb{E}_n \xrightarrow{W^1}_w \mathbb{B}_F$ if and only if $X \in \mathcal{L}^{4,2}$, where

$$W^1 := \left\{ f \in L^1 : \|f\|_{W^1} := \int_0^\infty (1+t)|f(t)| dt < \infty \right\}. \quad (23)$$

Our first task is to find processes expressed as continuous functionals of \mathbb{E}_n and equivalent to $\mathbb{X}_n(d)$. To start, for $t \geq 0$, we decompose $\mathbb{X}_n(d)$ in the following way:

$$\mathbb{X}_n(d) = \mathbb{A}_n(d) + \mathbb{B}_n(d) + \mathbb{C}_n(d), \quad (24)$$

where

$$\begin{aligned} \mathbb{A}_n(\omega) &:= \mathbb{E}_n, & \mathbb{B}_n(\omega) &:= \sqrt{n}(G_\mu - G_{\hat{\mu}}), & \mathbb{C}_n(\omega) &:= 0, \\ \mathbb{A}_n(\bar{\omega}) &:= \mathbb{E}_n/\hat{\mu}, & \mathbb{B}_n(\bar{\omega}) &:= \sqrt{n}(G_\mu - G_{\hat{\mu}})/\hat{\mu}, & \mathbb{C}_n(\bar{\omega}) &:= g_{\bar{\omega}}\sqrt{n}(\mu - \hat{\mu})/\hat{\mu}, \\ \mathbb{A}_n(\zeta_2, t) &:= \int_t^\infty \mathbb{E}_n, & \mathbb{B}_n(\zeta_2, t) &:= \sqrt{n} \int_t^\infty (G_\mu - G_{\hat{\mu}}), & \mathbb{C}_n(\zeta_2) &:= 0, \\ \mathbb{A}_n(\bar{\zeta}_2, t) &:= \frac{1}{\hat{\mu}^2} \int_t^\infty \mathbb{E}_n, & \mathbb{B}_n(\bar{\zeta}_2, t) &:= \frac{\sqrt{n}}{\hat{\mu}^2} \int_t^\infty (G_\mu - G_{\hat{\mu}}), & \mathbb{C}_n(\bar{\zeta}_2) &:= g_{\bar{\zeta}_2} \frac{\sqrt{n}(\mu^2 - \hat{\mu}^2)}{\hat{\mu}^2}. \end{aligned}$$

The following lemma provides equivalent expressions for the processes defined above.

Lemma 3. *Let X be a positive random variable with mean $\mu > 0$. For $t \geq 0$, the following assertions hold:*

- (a) If $X \in \mathcal{L}^{4/3}$, $\mathbb{B}_n(\omega) \stackrel{L^1}{\sim} \tilde{\mathbb{B}}_n(\omega)$, where $\tilde{\mathbb{B}}_n(\omega, t) := \sqrt{n}(\hat{\mu} - \mu)te^{-t/\mu}/\mu^2$.
- (b) If $X \in \mathcal{L}^{2,1}$, $\mathbb{A}_n(\bar{\omega}) \stackrel{L^1}{\sim} \tilde{\mathbb{A}}_n(\bar{\omega}) := \mathbb{E}_n/\mu$.
- (c) If $X \in \mathcal{L}^{4/3}$, $\mathbb{B}_n(\bar{\omega}) \stackrel{L^1}{\sim} \tilde{\mathbb{B}}_n(\bar{\omega})$, where $\tilde{\mathbb{B}}_n(\bar{\omega}, t) := \sqrt{n}(\hat{\mu} - \mu)te^{-t/\mu}/\mu^3$.
- (d) If $X \in \mathcal{L}^{4/3}$, $\mathbb{C}_n(\bar{\omega}) \stackrel{L^1}{\sim} \tilde{\mathbb{C}}_n(\bar{\omega}) := \sqrt{n}(\mu - \hat{\mu})g_{\bar{\omega}}/\mu$.
- (e) If $X \in \mathcal{L}^{4/3}$, $\mathbb{B}_n(\zeta_2) \stackrel{L^1}{\sim} \tilde{\mathbb{B}}_n(\zeta_2)$, where $\tilde{\mathbb{B}}_n(\zeta_2, t) := \sqrt{n}(\hat{\mu} - \mu)(1 + t/\mu)e^{-t/\mu}$.
- (f) If $X \in \mathcal{L}^{4,2}$, $\mathbb{A}_n(\bar{\zeta}_2) \stackrel{L^1}{\sim} \tilde{\mathbb{A}}_n(\bar{\zeta}_2)$, where $\tilde{\mathbb{A}}_n(\bar{\zeta}_2, t) := \int_t^\infty \mathbb{E}_n/\mu^2$.
- (g) If $X \in \mathcal{L}^{4/3}$, $\mathbb{B}_n(\bar{\zeta}_2) \stackrel{L^1}{\sim} \tilde{\mathbb{B}}_n(\bar{\zeta}_2)$, where $\tilde{\mathbb{B}}_n(\bar{\zeta}_2, t) := \sqrt{n}(\hat{\mu} - \mu)(1 + t/\mu)e^{-t/\mu}/\mu^2$.
- (h) If $X \in \mathcal{L}^2$, $\mathbb{C}_n(\bar{\zeta}_2) \stackrel{L^1}{\sim} \tilde{\mathbb{C}}_n(\bar{\zeta}_2) := \sqrt{n}(\mu - \hat{\mu})2g_{\bar{\zeta}_2}/\mu$.

Proof. To show part (a), we use the mean value theorem twice to obtain

$$\begin{aligned}\|\mathbb{B}_n(\omega) - \widetilde{\mathbb{B}}_n(\omega)\|_1 &= \sqrt{n}|\hat{\mu} - \mu| \int_0^\infty t |e^{-t/\mu_t}/\mu_t^2 - e^{-t/\mu}/\mu^2| dt \\ &\leq \sqrt{n}(\hat{\mu} - \mu)^2 \int_0^\infty t |2 - t/\xi_t| e^{-t/\xi_t}/\xi_t^3 dt,\end{aligned}\quad (25)$$

where μ_t and ξ_t are points between μ and $\hat{\mu}$. Finally, the last integral in (25) can be bounded by

$$\int_0^\infty t |2 - t/\xi_t| e^{-t/\xi_t}/\xi_t^3 dt \leq 2(\mu + \hat{\mu}) \frac{\max\{\mu, \hat{\mu}\}^2}{\min\{\mu, \hat{\mu}\}^4} \rightarrow 4/\mu \quad \text{a.s.} \quad (26)$$

Therefore, from (25)-(26), and by the Kolmogorov, Marcinkiewicz and Zygmund strong law of large numbers (see for instance [20, Theorem 3.23]), we see that, whenever $X \in \mathcal{L}^{4/3}$, $\|\mathbb{B}_n(\omega) - \widetilde{\mathbb{B}}_n(\omega)\|_1 \rightarrow 0$ a.s.

To see (b), we note that $\|\mathbb{A}_n(\bar{\omega}) - \widetilde{\mathbb{A}}_n(\bar{\omega})\|_1 = \|\mathbb{E}_n\|_1 |\mu - \hat{\mu}|/(\mu\hat{\mu})$. From Lemma 2 (a), we have that $\|\mathbb{E}_n\|_1 \rightarrow_d \|\mathbb{B}_F\|_1$ and the conclusion follows from the strong law of large numbers.

Part (c) follows from (a), as it is straightforward to check that $\mathbb{B}_n(\bar{\omega}) \stackrel{L^1}{\sim} \mathbb{B}_n(\omega)/\mu$, whenever $X \in \mathcal{L}^{4/3}$.

Part (d) is direct, whereas part (e) can be found in [4, Lemma 1]. The proof of part (f) is similar to the one for (b) by using Lemma 2 (b).

To show part (g), we first observe that, from part (e), we have that $\widetilde{\mathbb{B}}_n(\bar{\zeta}_2) \stackrel{L^1}{\sim} \mathbb{B}_n(\zeta_2)/\mu^2$. Further, some simple computations show that

$$\|\mathbb{B}_n(\bar{\zeta}_2) - \mathbb{B}_n(\zeta_2)/\mu^2\|_1 = \sqrt{n}(\mu - \hat{\mu})^2(1/\mu + 1/\hat{\mu})^2,$$

and the conclusion follows.

Finally, it can be checked that

$$\|\mathbb{C}_n(\bar{\zeta}_2) - \widetilde{\mathbb{C}}_n(\bar{\zeta}_2)\|_1 = \sqrt{n}(\mu - \hat{\mu})^2(1/\mu + 1/\hat{\mu}) \bar{\zeta}_2(X, Y_\mu).$$

As $\bar{\zeta}_2(X, Y_\mu) < \infty$ if and only if $X \in \mathcal{L}^2$, we conclude that (h) is fulfilled and the proof of the lemma is complete. \square

The next corollary, which is a consequence of Lemma 3 and (24), shows that $\mathbb{X}_n(d)$ are equivalent in L^1 to certain continuous functionals of the empirical process \mathbb{E}_n .

Corollary 3. *Let X be a positive random variable with mean $\mu > 0$. The following assertions hold:*

(i) *If $X \in \mathcal{L}^{4/3}$, we have that $\mathbb{X}_n(\omega) \stackrel{L^1}{\sim} \phi_\omega(\mathbb{E}_n)$, where $\phi_\omega : L^1 \rightarrow L^1$ is the linear operator defined by*

$$\phi_\omega(f, t) := f(t) - \frac{t}{\mu^2} e^{-t/\mu} \int_0^\infty f(x) dx, \quad t \geq 0.$$

Moreover, $\|\phi_\omega(f)\|_1 \leq 2\|f\|_1$, and ϕ_ω is therefore continuous.

(ii) If $X \in \mathcal{L}^{2,1}$, we have that $\mathbb{X}_n(\bar{\omega}) \stackrel{L^1}{\sim} \phi_{\bar{\omega}}(\mathbb{E}_n)$, where $\phi_{\bar{\omega}} : L^1 \rightarrow L^1$ is the linear operator defined by

$$\phi_{\bar{\omega}}(f, t) := \frac{1}{\mu} \left[f(t) + \left(g_{\bar{\omega}}(t) - \frac{t}{\mu^2} e^{-t/\mu} \right) \int_0^\infty f(x) dx \right], \quad t \geq 0.$$

Moreover, $\|\phi_{\bar{\omega}}(f)\|_1 \leq \|f\|_1(2 + \bar{\omega}(X, Y_\mu))/\mu$, and $\phi_{\bar{\omega}}$ is therefore continuous.

(iii) If $X \in \mathcal{L}^{4/3}$, we have that $\mathbb{X}_n(\zeta_2) \stackrel{L^1}{\sim} \phi_{\zeta_2}(\mathbb{E}_n)$, where $\phi_{\zeta_2} : W^1 \rightarrow L^1$ is the linear operator defined by

$$\phi_{\zeta_2}(f, t) := \int_t^\infty f(x) dx - \left(1 + \frac{t}{\mu} \right) e^{-t/\mu} \int_0^\infty f(x) dx, \quad t \geq 0.$$

Moreover, $\|\phi_{\zeta_2}(f)\|_1 \leq (1 + 2\mu) \|f\|_{W_1}$, and ϕ_{ζ_2} is therefore continuous.

(iv) If $X \in \mathcal{L}^{4,2}$, we have that $\mathbb{X}_n(\bar{\zeta}_2) \stackrel{L^1}{\sim} \phi_{\bar{\zeta}_2}(\mathbb{E}_n)$, where $\phi_{\bar{\zeta}_2} : W^1 \rightarrow L^1$ is the linear operator defined by

$$\phi_{\bar{\zeta}_2}(f, t) := \frac{1}{\mu^2} \left[\int_t^\infty f(x) dx + \left(2\mu g_{\bar{\zeta}_2}(t) - \left(1 + \frac{t}{\mu} \right) e^{-t/\mu} \right) \int_0^\infty f(x) dx \right], \quad t \geq 0.$$

Moreover, $\|\phi_{\bar{\zeta}_2}(f)\|_1 \leq \|f\|_{W_1} [1 + 2\mu(1 + \bar{\zeta}_2(X, Y_\mu))]/\mu^2$, and $\phi_{\bar{\zeta}_2}$ is therefore continuous.

We are now in condition to prove Theorems 1 and 2.

Proof of Theorem 1. Assume that (a) holds, i.e., $X \in \mathcal{L}^{2,1}$. For $d = \omega$ or $d = \bar{\omega}$, from Lemma 2 (a) and Corollary 3 (i) and (ii), we have that $\mathbb{X}_n(d) \stackrel{L^1}{\sim} \phi_d(\mathbb{E}_n) \xrightarrow{L^1}_{\mathbf{w}} \phi_d(\mathbb{B}_F) = \mathbb{X}_F(d)$, by the continuous mapping theorem. We conclude that (b) and (c) hold.

Conversely, let us assume first that (b) is satisfied. By Corollary 3 (i), if $X \in \mathcal{L}^{4/3}$, we obtain that $\phi_\omega(\mathbb{E}_n) \xrightarrow{L^1}_{\mathbf{w}} \mathbb{X}_F(\omega)$. Observe now that $\phi_\omega(\mathbb{E}_n)$ can be rewritten as a normalized sum in the following way

$$\phi_\omega(\mathbb{E}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{Y}_i(\omega),$$

where $\mathbb{Y}_1(\omega), \dots, \mathbb{Y}_n(\omega)$ are n independent copies of the zero mean process

$$\mathbb{Y}(\omega, t) := P(X > t) - I_{\{X > t\}} + (X - \mu)te^{-t/\mu}/\mu^2, \quad t \geq 0. \quad (27)$$

This means that the process $\mathbb{Y}(\omega)$ satisfies the CLT in L^1 (and implies that $\mathbb{X}_F(\omega)$ is a centered Gaussian process), which is equivalent (see [2, Exercise 14, p. 205]) to

$$\int_0^\infty \sqrt{E\mathbb{Y}(\omega, t)^2} dt < \infty. \quad (28)$$

In particular, this implies that $X \in \mathcal{L}^2$ and denoting $\mathbb{Z}(t) := P(X > t) - I_{\{X > t\}}$ ($t \geq 0$), from (27), (28), and by Minkowski inequality, we have that

$$\int_0^\infty \sqrt{E\mathbb{Z}(t)^2} dt \leq \sigma + \int_0^\infty \sqrt{E\mathbb{Y}(\omega, t)^2} dt < \infty,$$

where σ is the standard deviation of X . Last inequality amounts to $X \in \mathcal{L}^{2,1}$ and the proof is complete.

For the proof that (c) implies (a), it is enough to note that $\mathbb{X}_n(\bar{\omega}) \xrightarrow{L^1} \mathbb{X}_F(\bar{\omega})$ is equivalent to $\hat{\mu} \mathbb{X}_n(\bar{\omega}) \xrightarrow{L^1} \mu \mathbb{X}_F(\bar{\omega})$. The rest of the proof runs as with $d = \omega$ \square

Proof of Theorem 2. To show that part (a) implies (b) and (c) it is enough to follow the same steps as in the proof of the same implications in Theorem 1. We omit the details.

To finish, we will show that part (c) implies (a) (the remaining implication “(b) \Rightarrow (a)” is simpler and similar). Let us assume that (c) is satisfied. In this situation, it is clear that $X \in \mathcal{L}^2$, as this integrability condition amounts to saying that the process $\mathbb{X}(\bar{\zeta}_2)$ has its paths in L^1 a.s. We have that $\hat{\mu}^2 \mathbb{X}_n(\bar{\zeta}_2) \xrightarrow{L^1} \mu^2 \mathbb{X}_F(\bar{\zeta}_2)$. Further, by Lemma 3, we conclude that $\hat{\mu}^2 \mathbb{X}_n(\bar{\zeta}_2) \stackrel{L^1}{\sim} \bar{\mathbb{Z}}_n$, where

$$\bar{\mathbb{Z}}_n(t) := \int_t^\infty \mathbb{E}_n + h(t)\sqrt{n}(\mu - \hat{\mu}), \quad t \geq 0,$$

with

$$h(t) := 2\mu g_{\bar{\zeta}_2}(t) - (1 + t/\mu)e^{-t/\mu}, \quad t \geq 0. \quad (29)$$

Some computations show that $\bar{\mathbb{Z}}_n$ can be rewritten as a normalized sum in the following way:

$$\bar{\mathbb{Z}}_n = \frac{1}{\sqrt{n}} \sum_{i=1}^n \mathbb{Z}_i(\bar{\zeta}_2),$$

where $\mathbb{Z}_1(\bar{\zeta}_2), \dots, \mathbb{Z}_n(\bar{\zeta}_2)$ are n independent copies of the zero mean process

$$\mathbb{Z}(\bar{\zeta}_2, t) := E(X - t)_+ - (X - t)_+ + h(t)(\mu - X).$$

Therefore, the process $\mathbb{Z}(\bar{\zeta}_2)$ satisfies the CLT in L^1 (and we see that $\mathbb{X}_F(\bar{\zeta}_2)$ is a centered Gaussian process). Using again [2, Exercise 14, p. 205], we obtain that

$$\int_0^\infty \sqrt{E\mathbb{Z}(\bar{\zeta}_2, t)^2} dt < \infty. \quad (30)$$

Finally, by Minkowski inequality and Fubini theorem, we have that

$$\int_0^\infty \sqrt{E(X - t)_+^2} dt \leq EX^2/2 + \sigma \|h\|_1 + \int_0^\infty \sqrt{E\mathbb{Z}(\bar{\zeta}_2, t)^2} dt.$$

As $\|h\|_1 \leq 2\mu(1 + \bar{\zeta}_2(X, Y_\mu)) < \infty$, from (30), we conclude that $\int_0^\infty \sqrt{E(X - t)_+^2} dt < \infty$. This implies that $X \in \mathcal{L}^{4,2}$ as $t^2 P(X > 2t) \leq E(X - t)_+^2$. \square

Proof of Lemma 1. First, we note that we can assume that $g \geq 0$. If this is not the case, it is enough to write $g = |g| \operatorname{sgn}(g)$. In such a case, we have that

$$\begin{aligned} \rho_n(f, g) &= \int_{I_g} |f| + \int_{I_g^c} (|f + \sqrt{n}g| - \sqrt{n}g) \\ &= \int_{I_g} |f| + \int_{I_g^c \cap \{f + \sqrt{n}g \geq 0\}} f - \int_{I_g^c \cap \{f + \sqrt{n}g < 0\}} (f + 2\sqrt{n}g). \end{aligned} \quad (31)$$

By the dominated convergence theorem, the second integral in (31) converges to $\int_{I_g^c} f$. For the last integral in (31), we have that

$$\begin{aligned} \int_{I_g^c \cap \{f + \sqrt{n}g < 0\}} |f + 2\sqrt{n}g| &\leq 2 \int_{I_g^c \cap \{f + \sqrt{n}g < 0\}} |f + \sqrt{n}g| + \int_{I_g^c \cap \{f + \sqrt{n}g < 0\}} |f| \\ &\leq 3 \int_{I_g^c \cap \{f + \sqrt{n}g < 0\}} |f|. \end{aligned} \quad (32)$$

Again, by the dominated convergence theorem we conclude that the integral in (32) goes to 0, and hence we conclude that $\lim_{n \rightarrow \infty} \rho_n(f, g) = \rho(f, g)$.

Finally, if $f_n \rightarrow f$ in L^1 , by Minkowski inequality, we obtain that

$$\begin{aligned} |\rho_n(f_n, g) - \rho(f, g)| &\leq |\rho_n(f_n, g) - \rho_n(f, g)| + |\rho_n(f, g) - \rho(f, g)| \\ &\leq \|f_n - f\|_1 + |\rho_n(f, g) - \rho(f, g)|. \end{aligned}$$

Therefore, we see that $\lim_{n \rightarrow \infty} \rho_n(f_n, g) = \rho(f, g)$, and the proof is complete. \square

Proof of Theorem 3. We have that $\delta_n(d, F) = \rho_n(\mathbb{X}_n(d), g_d)$, with ρ_n defined in (14). The conclusion follows from Theorems 1 and 2, Lemma 1, and an extended version of the continuous mapping theorem (see [34, Theorem 18.11]). \square

Proof of Corollary 2. Under the assumptions of this corollary, from Theorem 3, we have that

$$\rho(\mathbb{X}_F(d), g_d) = \int_0^\infty \mathbb{X}_F(d, t) \operatorname{sgn}(g_d(t)) dt.$$

As $\mathbb{X}_F(d)$ is a centered Gaussian process and g_d is nonrandom, we conclude that $\rho(\mathbb{X}_F(d), g_d)$ is normally distributed. \square

Acknowledgements

Research by A.B. and J.C. was supported by the Spanish MCyT grant MTM2013-44045-P. Research by J.C. was also supported by the Spanish MCyT grant MTM2011-27248. K.V.G. acknowledges the support from the Chandra ACIS Team contract SV4-74018 (G. Garmire & L. Townsley, PIs), issued by the Chandra X-ray Center, which is operated by the Smithsonian Astrophysical Observatory for and on behalf of NASA under contract NAS8-03060.

References

- [1] Ahsanullah, M. and Hamedani, G.G. (2010). *Exponential Distribution: Theory, Methods*. Nova Science Publishers, Inc.
- [2] Araujo, A. and Giné, E. (1980). *The Central Limit Theorem for Real and Banach Valued Random Variables*. Wiley.
- [3] Azlarov, T.A. and Volodin, N.A. (1986). *Characterization Problems Associated with the Exponential Distribution*. Springer.
- [4] Baíllo, A., Cárcamo, J. and Nieto, E. (2015). A test for convex dominance with respect to the exponential class based on an L^1 distance. *IEEE Transactions on Reliability*, 64, 71–82.
- [5] Balakrishnan, K. and Basu, A.P. (1995). *Exponential Distribution: Theory, Methods and Applications*. CRC Press.
- [6] Barlow, R.E. and Proschan, F. (1975). *Statistical Theory of Reliability and Life Testing: Probability Models*. Holt, Rinehart and Winston.
- [7] del Barrio, E., Giné, E. and Matrán, C. (1999). Central limit theorems for the Wasserstein distance between the empirical and the true distributions. *The Annals of Probability*, 27, 1009–1071.
- [8] Boutsikas, M.V. and Vaggelatou, E. (2002). On the distance between convex-ordered random variables, with applications. *Advances in Applied Probability*, 34, 349–374.
- [9] Byczkowski, T. (1977). Gaussian measures on L^p spaces $0 \leq p < \infty$. *Studia Mathematica*, 59, 249–261.
- [10] Denuit, M., Dhaene, J., Goovaerts, M. and Kaas, R. (2005). *Actuarial Theory for Dependent Risks: Measures, Orders and Models*. Wiley.
- [11] Feigelson, E.D., Getman, K., Townsley, L., Garmire, G., Preibisch, T., Grosso, N., Montmerle, T., Muench, A. and McCaughrean, M. (2005). Global X-ray properties of the Orion Nebula region. *The Astrophysical Journal Supplement Series*, 160, 379–389.
- [12] Grafakos, L. (2008). *Classical Fourier Analysis*. Springer.
- [13] Galambos, J. and Kotz, S. (1978). *Characterization of Probability Distributions*. Lecture Notes in Mathematics, Vol. 675. Springer.
- [14] Getman, K.V., Feigelson, E.D., Broos, P.S., Micela, G. and Garmire, G.P. (2008a). X-ray flares in Orion young stars. I. Flare characteristics. *The Astrophysical Journal*, 688, 418–436.
- [15] Getman, K.V., Feigelson, E.D., Micela, G., Jardine, M.M., Gregory, S.G. and Garmire, G.P. (2008b). X-ray flares in Orion young stars. II. Flares, magnetospheres, and protoplanetary disks. *The Astrophysical Journal*, 688, 437–455.

- [16] Getman, K.V., Flaccomio, E., Broos, P.S., Grosso, N., Tsujimoto, M., Townsley, L., Garmire, G.P., Kastner, J., Li, J., Harnden, F.R., Wolk, S., Murray, S.S., Lada, C.J., Muench, A.A., McCaughrean, M.J., Meeus, G., Damiani, F., Micela, G., Sciortino, S., Bally, J., Hillenbrand, L.A., Herbst, W., Preibisch, T. and Feigelson, E. (2005). Chandra Orion Ultradeep Project: Observations and source lists. *The Astrophysical Journal Supplement Series*, 160, 319–352.
- [17] Gibbs, A.L. and Su, F.E. (2002). On choosing and bounding probability metrics. *International Statistical Review*, 70, 419–435.
- [18] Huber, P.J. (1981). *Robust Statistics*. Wiley.
- [19] Johnson, N.J., Kotz, S. and Balakrishnan, N. (1994). *Continuous Univariate Distributions*. Volume 1. Second Edition. Wiley.
- [20] Kallemberg, O. (1997). *Foundations of Modern Probability*. Springer.
- [21] Lai, C.-D. and Xie, M. (2006). *Stochastic Ageing and Dependence for Reliability*. Springer.
- [22] Ledoux, M. and Talagrand, M. (2002). *Probability in Banach Spaces*. Springer.
- [23] Levy, H. (2006). *Stochastic Dominance: Investment Decision Making under Uncertainty*. 2nd edition. Springer.
- [24] Marshall, A.W. and Olkin, I. (2007). *Life Distributions*. Springer.
- [25] Neininger, R. and Rüschendorf, L. (2004a). A general limit theorem for recursive algorithms and combinatorial structures. *The Annals of Applied Probability*, 14, 378–418.
- [26] Neininger, R. and Rüschendorf, L. (2004b). On the contraction method with degenerate limit equation. *The Annals of Probability*, 32, 2838–2856.
- [27] Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer.
- [28] Rachev, S.T., Klebanov, L., Stoyanov, S.V. and Fabozzi, F. (2013). *The Methods of Distances in the Theory of Probability and Statistics*. Springer.
- [29] Rao, M.M. (1997). *Real and Stochastic Analysis: Recent Advances*. CRC Press.
- [30] Rubner, Y., Tomasi, C. and Guibas, L.J. (2000). The Earth Mover’s distance as a metric for image retrieval. *International Journal of Computer Vision*, 40, 2, 99–121.
- [31] Schulz, N.S. (2012). *The Formation and Early Evolution of Stars: From Dust to Stars and Planets*. Second edition. Springer.
- [32] Shaked, M. and Shanthikumar, J.G. (2006). *Stochastic Orders*. Springer.
- [33] Sriperumbudur, B.K., Fukumizu, K., Gretton, A., Scholkopf, B. and Lanckriet, G.R.G. (2012). On the empirical estimation of integral probability metrics. *Electronic Journal of Statistics*, 6, 1550–1599.

- [34] van der Vaart, A. (1998). *Asymptotic Statistics*. Cambridge University Press.
- [35] Zolotarev, V.M. (1997). *Modern Theory of Summation of Random Variables*. De Gruyter.

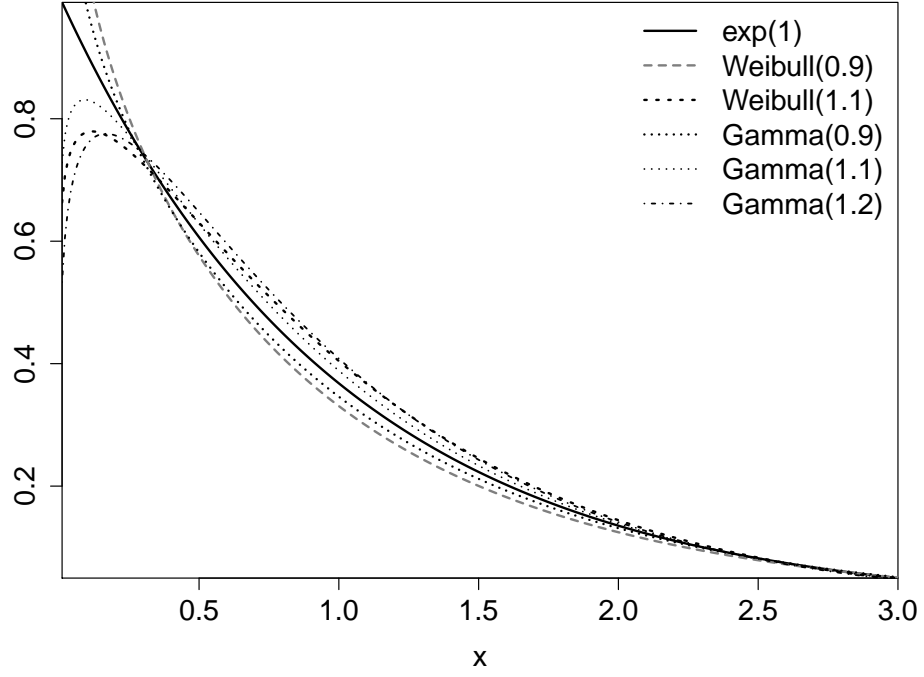


Figure 1: Probability densities used in the simulation study of Section 4.

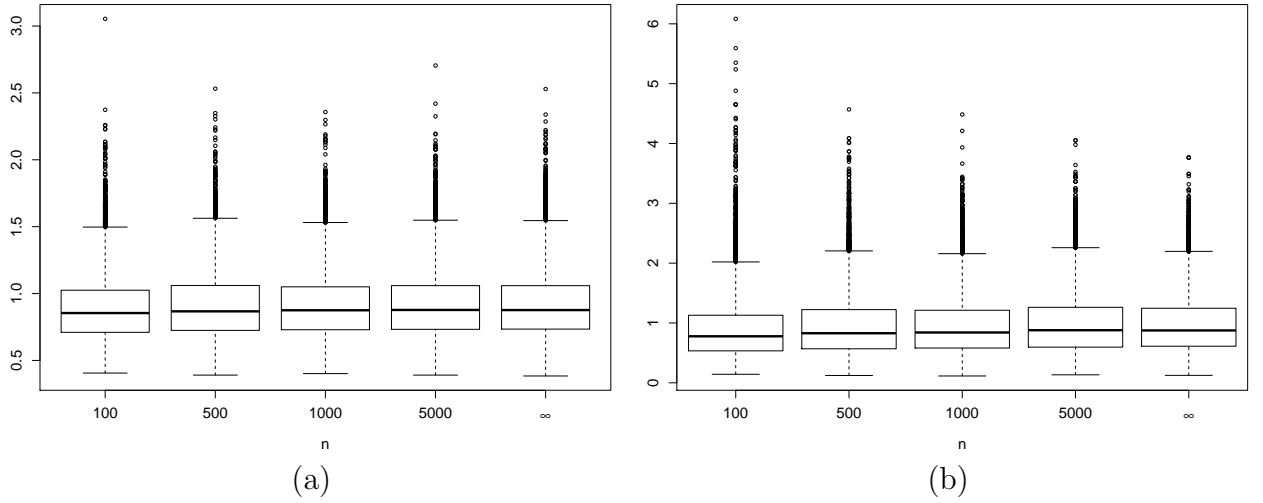


Figure 2: Boxplots of the statistic $\delta_n(d, F)$ and its asymptotic distribution for (a) $d = \bar{\omega}$ and (b) $d = \zeta_2$. The distribution F is exponential(1).

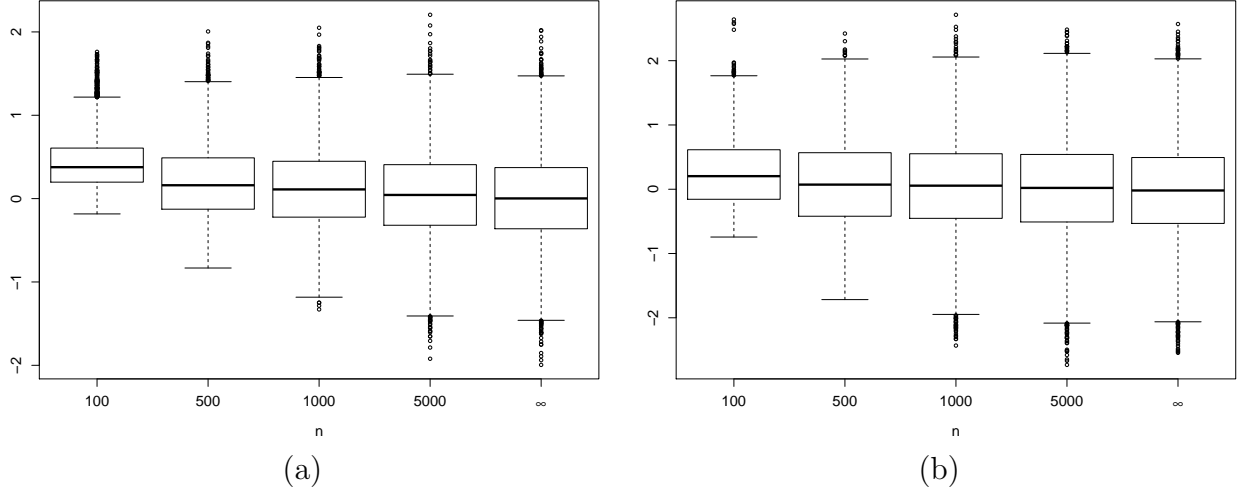


Figure 3: Boxplots of the statistic $\delta_n(d, F)$ and its asymptotic distribution for (a) $d = \bar{\omega}$ and (b) $d = \zeta_2$. The distribution F is Weibull with shape parameter $a = 1.1$.

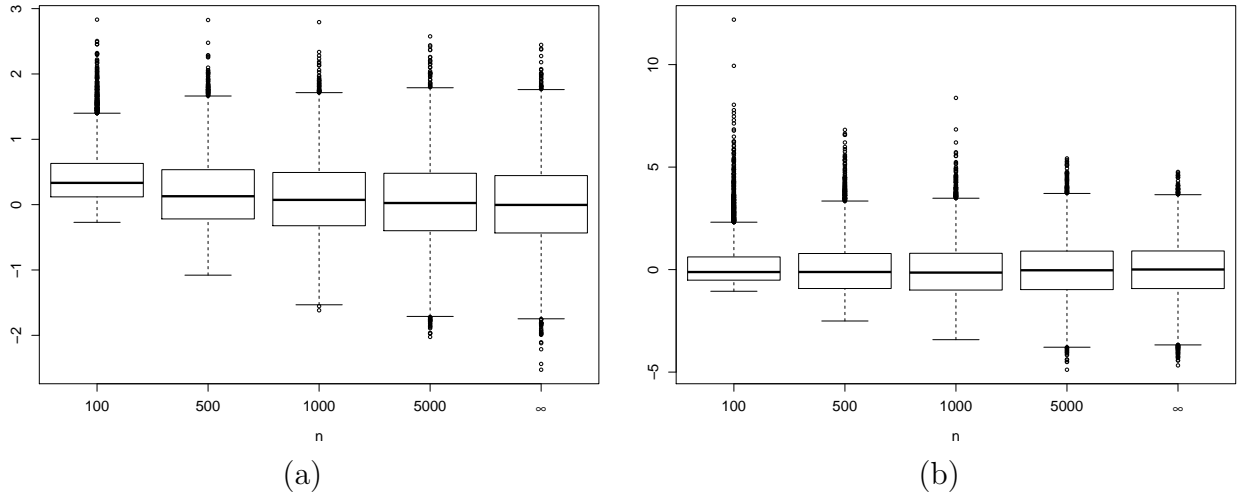


Figure 4: Boxplots of the statistic $\delta_n(d, F)$ and its asymptotic distribution for (a) $d = \bar{\omega}$ and (b) $d = \zeta_2$. The distribution F is Weibull with shape parameter $a = 0.9$.

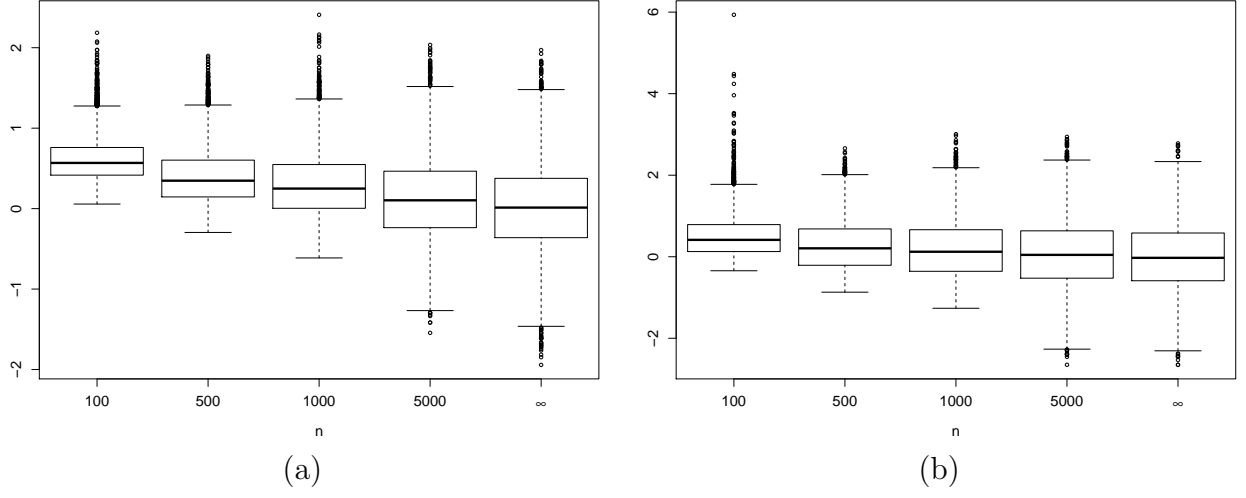


Figure 5: Boxplots of the statistic $\delta_n(d, F)$ and its asymptotic distribution for (a) $d = \bar{\omega}$ and (b) $d = \bar{\zeta}_2$. The distribution F is gamma with shape parameter $a = 1.1$.

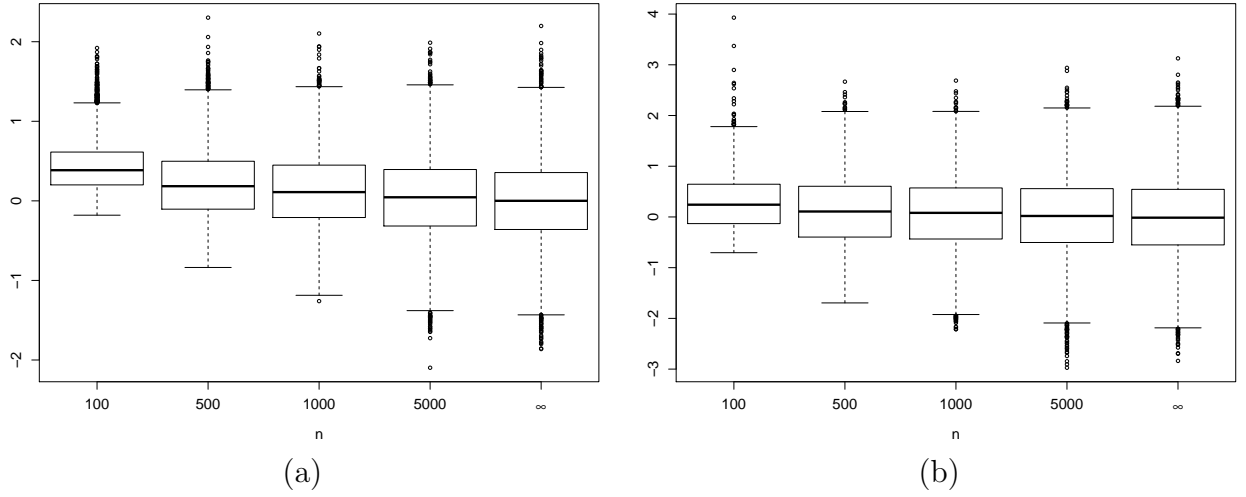


Figure 6: Boxplots of the statistic $\delta_n(d, F)$ and its asymptotic distribution for (a) $d = \bar{\omega}$ and (b) $d = \bar{\zeta}_2$. The distribution F is gamma with shape parameter $a = 1.2$.

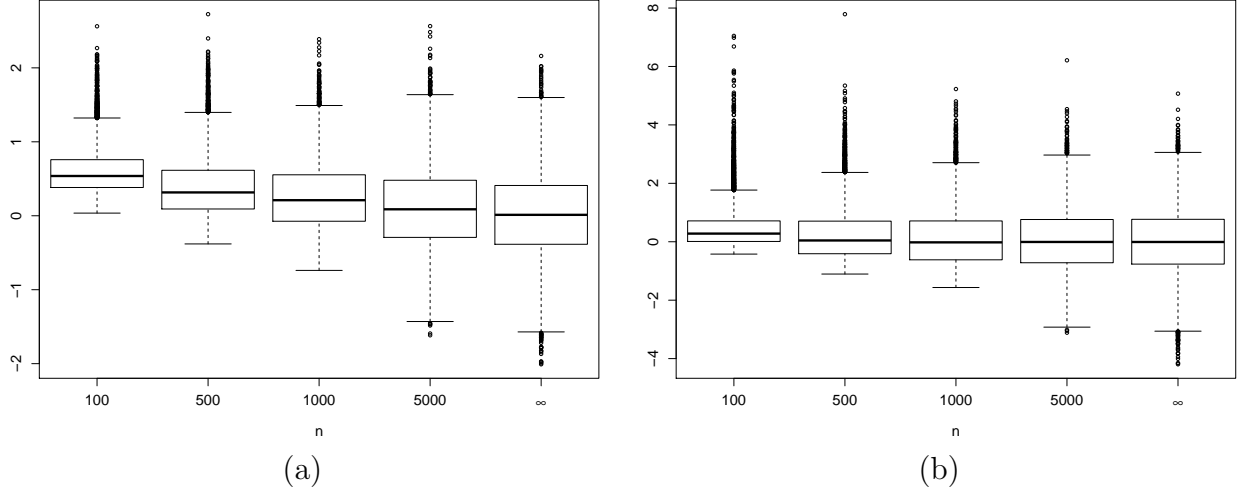


Figure 7: Boxplots of the statistic $\delta_n(d, F)$ and its asymptotic distribution for (a) $d = \bar{\omega}$ and (b) $d = \bar{\zeta}_2$. The distribution F is gamma with shape parameter $a = 0.9$.

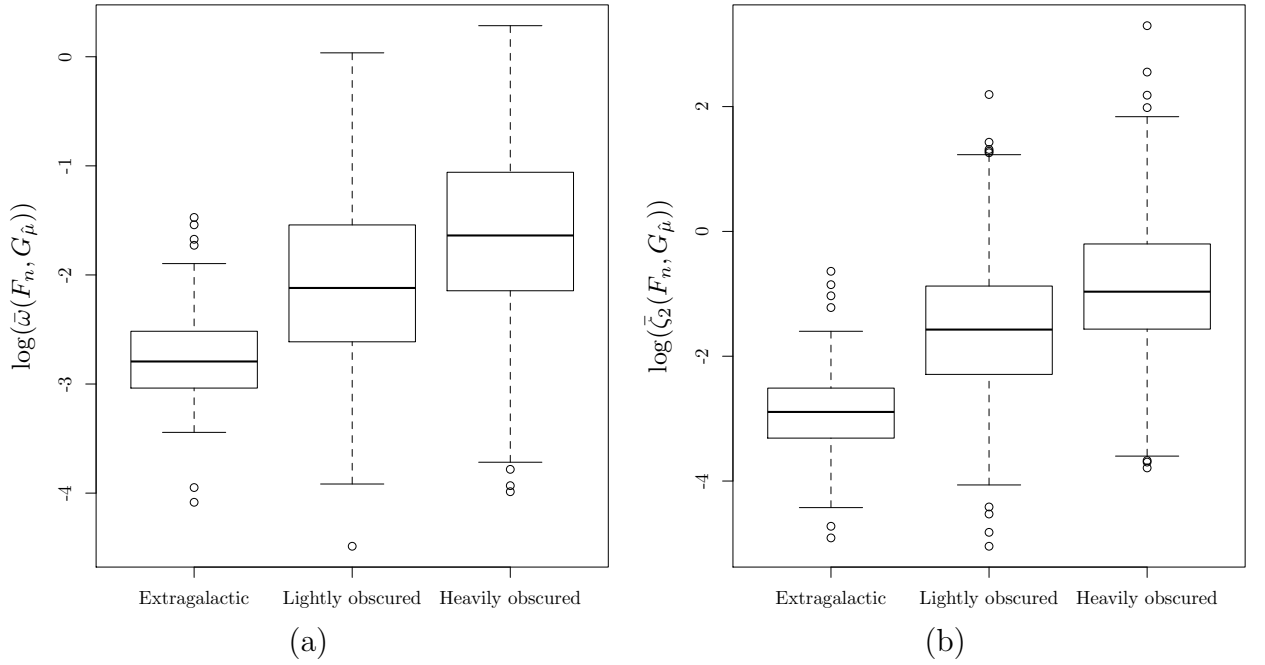


Figure 8: Analysis of COUP data. Boxplots of $\log(d(F_n, G_{\hat{\mu}}))$ for (a) $d = \bar{\omega}$ and (b) $d = \bar{\zeta}_2$.

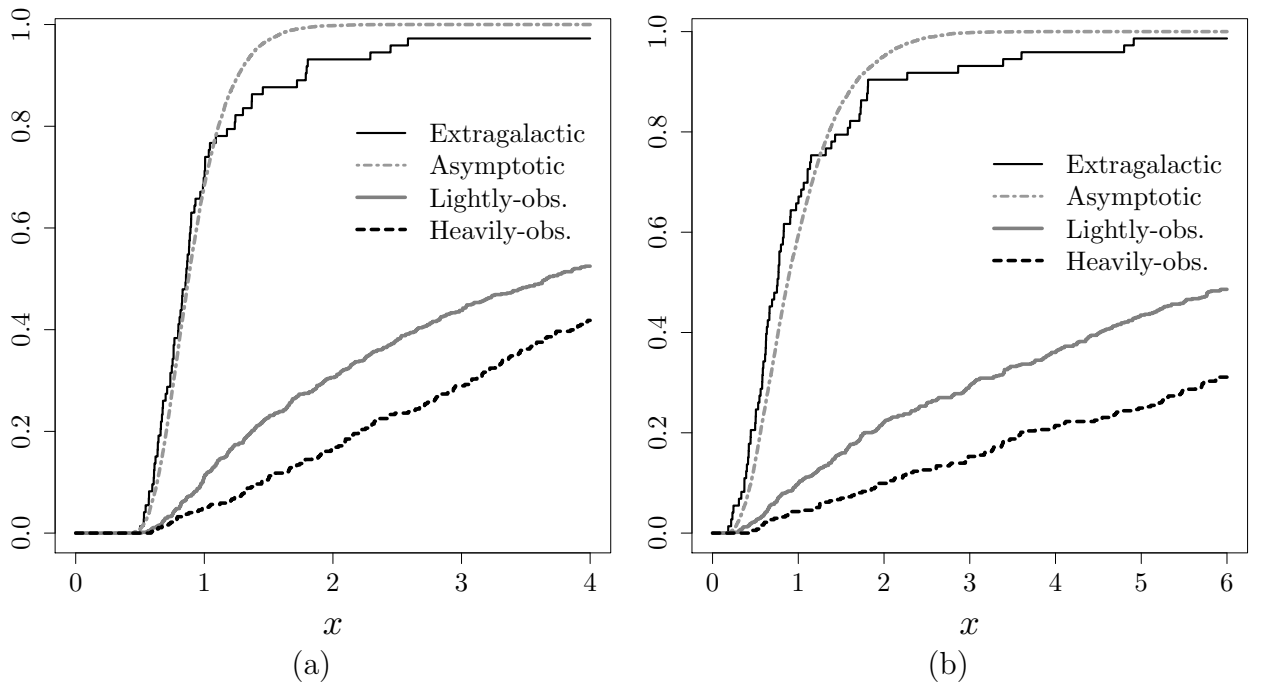


Figure 9: Analysis of COUP data. Asymptotic distribution function of $\delta_n(d, G_1)$ and empirical distribution functions of $\sqrt{n} d(F_n, G_{\hat{\mu}})$ for (a) $d = \bar{\omega}$ and (b) $d = \bar{\zeta}_2$.